On the Identification and Annotation of Emotional Properties of Verbs

Nikolaos Papatheodorou[1], Pepi Stavropoulou[1], Dimitrios Tsonos[1],
Georgios Kouroupetroglou[1], Dimitris Spiliotopoulos[1],
and Charalambos Papageorgiou[2]

[1] Department of Informatics and Telecommunications,
National and Kapodistrian University of Athens, Greece
koupe@di.uoa.gr
[2] Department of Psychiatry, Eginition Hospital,
National and Kapodistrian University of Athens, Greece

Abstract. Adequate and reliable lexical resources are essential for effective sentiment analysis and opinion mining. This paper proposes a methodology for the emotional assessment and annotation of words. The process is based on the Self Assessment Manikin test, and is coupled with two psychometric measurements for identifying possible bias due to the annotator's psychological condition and personality: the EPQ scale and the SCL-90-R scale. A web based tool was developed to support the process. The methodology was validated through a pilot study in which 10 participants were asked to assess the emotional state elicited by each of 75 verbs that were used as stimuli. Results are compared with SentiWordNet's emotional scoring on respective verbs, and primarily show logical continuity and consistency.

Keywords: Verbs, Emotional State, SentiWordNet.

1 Introduction

Sentiment analysis and opinion mining [1] have become an important area of research pertaining to a range of applications such as socially aware conversation agents, affective Text to Speech engines, speech analytics solutions and other data mining applications that allow for quality monitoring of people's views on products, services and processes. The majority of machine learning techniques and algorithms for opinion mining depend upon the presence of negative and positive words, which are typically defined in resources such as sentiment lexicons. Accordingly, development of such adequate and reliable resources is crucial for effective sentiment analysis. Senti-WordNet [2, 3, 4], for example, is an attempt to extend WordNet [5] with sentiment scores for each sense in the semantic network, in order to aid tasks such as affective reasoning, subject engagement in e-learning, polarity analysis and computational humor.

While the SentiWordNet labeling handles polarity identification, a richer sentiment analysis on the emotional level is often needed for affective computing tasks. This paper presents a novel methodology for the effective classification and annotation of

the emotional properties of words and phrases. The methodology supports an emotional assessment procedure based on the Self Assessment Manikin (SAM) Test [6, 7, 8], measuring emotions elicited by words on three dimensions: pleasure, arousal and dominance. This approach is evaluated through a dedicated web-based environment, developed for this task, so that it can be easily and immediately accessible to the public, potentially ultimately serving as a crowdsourcing technique for fast and cost effective creation of lexical resources [9]. The process is coupled with psychometric questionnaires (EPQ and SCL-90-R scale), in order to verify the reliability of the annotation and prevent any bias due to the psychological condition of the annotator. In addition, the results of this process can provide insight into how people respond emotionally to the words they use, listen and/or read, as well as the affective meaning of the verb they perceive.

We further review existing resources, SentiWordNet in particular, and present the results of a pilot experiment run to evaluate the validity of the proposed emotional assessment methodology by comparing the annotation outcome of the procedure to SentiWordNet sentiment scores. For evaluation, this work focuses on Greek verbs. However, the methodology is also applicable to other languages, such as English, which can provide even more data that can lead to a more comprehensive survey with many ramifications. The techniques presented and evaluated in this paper would be useful for all natural language processing research where emotion-level sentiment analysis is required such as emotion prediction [10]. The main contribution of this work is a psychology-based method for identification of emotional properties of verbs based on pleasure, arousal, dominance parameters and comparative evaluation of the SentiWordNet polarity values.

In the following sections, we first present SentiWordNet and other lexical resources that were utilized for the collection of materials used in the pilot study. In section 3 we present the emotional assessment process. Finally, the results of the study are presented and discussed.

2   Lexical Resources for Sentiment Analysis

2.1   SentiWordNet

SentiWordNet constitutes an extension to the original WordNet developed to support "sentiment classification and opinion mining applications" [4]. In SentiWordNet each synset of WordNet (i.e. each set of (near-) synonyms corresponding to the same sense) is assigned a numerical score corresponding to the polarity of the synset, i.e. how objective, positive or negative the sense is (and consequently how objective, positive or negative each word within the synset is). More specifically, each synset of WordNet is associated to three numerical scores Obj(s), Pos(s) and Neg(s), describing the degree of Objectivity, Positivity and Negativity of each term contained in the synset. Each of the three scores ranges from 0.0 to 1.0, and their sum adds up to 1.0 (Figure 1). Essentially, this means that a synset may have nonzero scores for all three polarity axes, which would indicate that the corresponding terms have each of the three opinion-related properties only to a certain degree.

A semi-supervised algorithm was used for the annotation of the complete set of WordNet synsets with polarity (objective, positive and negative) scores. Initially, a small set of "seed" synsets was used derived from sentiment lexicons. The set was then expanded based on the semantic relations (antonymy, hyponymy etc.) between synsets to be used as a training corpus for the polarity classifiers. The final classification algorithm was based on the presence or/and absence of negative, objective and positive senses within each synset's gloss. Due to the fact that scores are automatically assigned, and contextual information is not taken into account, mistakes may sometimes occur in the classification. For example, protect#1 (gloss: "shield from danger, injury, destruction, or damage") assigns a 0.75 score on the negative axis and a 0.25 score on the objective axis, arguably because of the presence of many negative terms in the synset's gloss. Not taking context into account, the classification algorithm fails to acknowledge the full impact of the term "shield". To minimize the possibility of mistakes in the classification of SentiWordNet verbs used for the experimental evaluation at hand, we only used verbs that scored over 0.8 on one of the three axes.
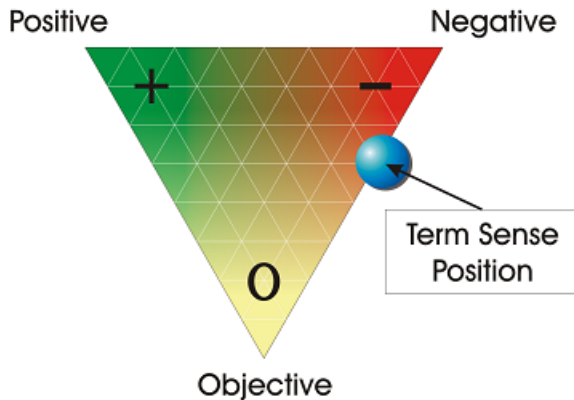

Fig. 1. Graphical model representation of the three scores for each synset

2.2   Corpus of Contemporary American English / Stimuli Selection

In order to reduce the large number of verbs available for the experimental evaluation, we referred to the lexical database of the Corpus of Contemporary American English [11], which is claimed to contain some of the most accurate lemma frequency lists. We first identified the 1000 most frequently used verbs within the list of the 5000 most frequently used lemmas. Next we extracted the polarity scoring of these verbs in SentiWordNet. However, a significant percentage of these verbs rated similarly on more than one axis (<0.6 per axis), which practically made it impossible to unambiguously determine their polarity. Therefore, in order to have a clearer assessment of the emotional value of verbs, we only used the verbs that scored high (>0.8) on the positive, negative or neutral axis, and were thus assigned with an unambiguous emotional value. As a result, we ended up with a total of 75 verbs, 25 in each category

(positive, negative, and neutral). These verbs were then translated to Greek. The translation was done in accordance with the widely known online dictionary Babylon English - Greek [12]. Table 1 shows the 75 verbs classified according to their basic emotional value (Positive, Negative, Neutral).

Table 1. The 75 verbs classified according to their SentiWordNet scoring

Verb ID	Positive		Neutral		Negative	
	Word	Greek Translation	Word	Greek Translation	Word	Greek Translation
1	prefer	προτιμώ	do	κάνω	deny	αρνούμαι
2	fit	ταιριάζω	say	λέγω	protect	προστατεύω
3	question	ερωτώ	go	πηγαίνω	hate	μισώ
4	deserve	αξίζω	can	μπορώ	complain	παραπονούμαι
5	suit	αρμόζω	make	κατασκευάζω	damage	ζημιώνω
6	accommodate	διευκολύνω	know	γνωρίζω	abuse	καταχρώμαι
7	qualify	δικαιούμαι	will	διαθέτω	fear	φοβούμαι
8	admire	θαυμάζω	think	σκέφτομαι	murder	δολοφονώ
9	please	ευχαριστώ	take	λαμβάνω	average	υπολογίζω
10	donate	δωρίζω	see	βλέπω	apologize	απολογούμαι
11	love	αγαπώ	come	έρχομαι	spare	εξοικονομώ
12	teach	διδάσκω	look	κοιτάζω	lose	χάνω
13	save	σώζω	use	χρησιμοποιώ	face	αντικρίζω
14	check	ελέγχω	find	βρίσκω	imagine	φαντάζομαι
15	tend	τείνω	tell	διηγούμαι	worry	ανησυχώ
16	define	ορίζω	work	εργάζομαι	replace	αντικαθιστώ
17	contribute	συνεισφέρω	try	προσπαθώ	ignore	αγνοώ
18	respect	σέβομαι	ask	ζητώ	disappear	χάνομαι
19	possess	κατέχω	become	γίνομαι	warn	προειδοποιώ
20	satisfy	ικανοποιώ	leave	φεύγω	steal	κλέβω
21	rid	απαλλάσω	put	θέτω	mind	νοιάζομαι
22	lend	δανείζω	mean	εννοώ	confront	αντιμετωπίζω
23	rescue	διασώζω	keep	κρατώ	endure	αντέχω
24	diagnose	διαγιγνώσκω	let	αφήνω	injure	τραυματίζω
25	instruct	Καθοδηγώ	begin	αρχίζω	vanish	Εξαφανίζομαι

Analysis of the final set of verbs indicates the following: most positive and negative verbs do not directly refer to emotional states (e.g. "fear"), but rather refer to contexts that elicit positive or negative emotional responses (e.g. "murder" or

"rescue"). Accordingly, a lot of emotion state denoting verbs that have been typically used in psychological research on emotions [13] did not end up in this set due to their low frequency of occurrence (e.g. "resent") or their ambiguous scoring in Senti-WordNet (e.g., "satisfy", "encourage"). Finally, analysis of subcategorization and thematic roles associated with each verb showed that "experiencer" is a more frequent role for positive and negative terms compared to neutral (36%, 32% and 20% respectively), while more positive and negative verbs impose a [+human] selectional restriction on their subject/object compared to neutral ones (40% and 16% respectively).

Both observations are in line with long identified characteristics of emotion verbs and verbs of psychological state [14, 15]. Overall, the most frequent role in the thematic grid was that of "agent/actor".

3   Emotional Assessment Tool and Experimental Process

The tool developed for the estimation of the emotions elicited by each verb is based on the Self Assessment Manikin (SAM) Test [6, 7, 8] and the "Pleasure", "Arousal" and "Dominance" (PAD) three-dimensional model for measuring the effect [16]. The SAM method measures the emotional response, based on the dimensional model of emotions, allowing the users to self-assess and express their emotional response visually rather than verbally. The latter makes it fast and easy to use in the application at hand and further enables a more coherent, consistent, independent of cultural and linguistic characteristics assessment, suitable for use in different countries and cultural groups [7, 8].

SAM Test measures emotion scales on the PAD dimensions:

- "Pleasure" (also referred to as "evaluation", "valence"). It measures how pleasant, positive, negative or neutral, an emotion is. This dimension is primarily depicted on SentiWordNet's scoring.
- "Arousal" (also referred to as "activation", "activity"). It measures the intensity, strength of the emotion (e.g. frighten vs. terrify).
- "Dominance" (also referred to as "power", "potency"). It measures how controlling/dominant or submissive the emotion is (e.g. fear vs. anger).

Participants in the experimental procedure can choose one from at least five figures. In the present study we used a 5-point scale (Figure 2). To assess the emotional state of "Pleasure", the extreme values are a smiling and a sad figure. In case of "Arousal" one pole is represented by a figure of great vigor and the opposite by a calm manikin, with eyes closed. Similarly, in the "Dominance" dimension, submissiveness is represented as controlled by a small manikin and dominance by a large one. When evaluating the results, users' answers can be easily converted from scale points to a group of values which ranges in the interval [-1, 1] or [-100%, 100%]. The value "0" represents the neutral state in each of the three dimensions.
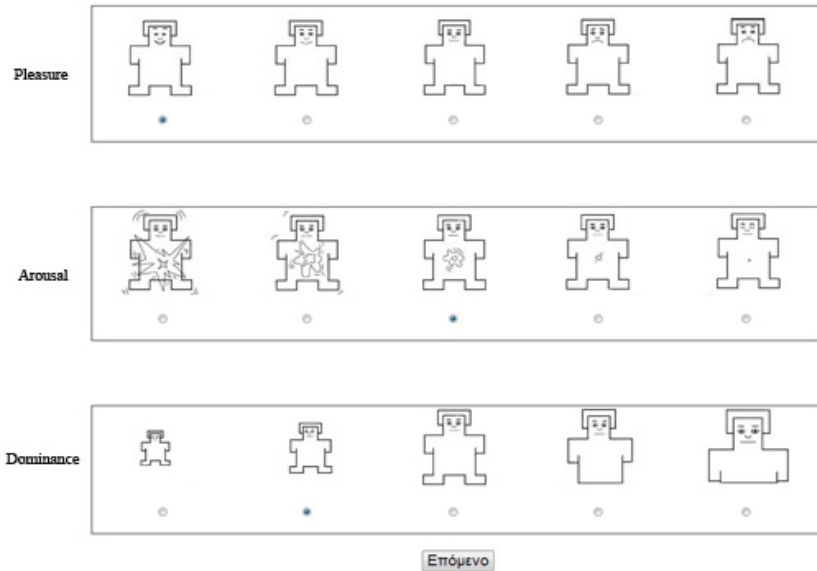
Fig. 2. The manikins of the 5-point scale SAM Test as presented during the test. The verbal expressions of "Pleasure", "Arousal" and "Dominance" do not appear during the test.

The implementation of the tool is based on PHP [17] and MySQL [18] in order to give the participants the possibility to automatically submit their answers from their own computer. Also, it gives researchers the functionality to easily and quickly retrieve and process the data. In addition, technologies like HTML [19], JavaScript [20] and CSS [21] were used in order to enhance the usability, administration and visual characteristics of the interface (e.g. to accurately control the projection time of the verb).

The second part of the experiment involved the completion of two questionnaires that evaluate the mental condition of the participant, specifically Scale EPQ and Scale SCL-90-R [22, 23]. The EPQ scale defines the personality in Eysenck's model and the SCL-90-R scale shows the actual self assessment psychological condition.

4   Results

Ten native Greek speakers participated in the experiment, 5 men and 5 women (mean age = 37,5 years, SD = 13,7 years). The stimuli were 75 Greek verbs, 25 positively, 25 negatively and 25 objectively classified (Table 1). Each participant was asked to fill in her/his demographic information and complete a consent form that she/he agrees to participate in the experiment. Then, they were familiarized with the annotation tool through a demo of the SAM test (three stimuli). After the completion of the experiment, they filled in the EPQ and SCL-90-R questionnaires. The mean duration of the total experimental procedure (introduction and familiarization with the tool, SAM test and psychometric tests) was 48.4 minutes.

Figures 3, 4 and 5 present the average values of participants' responses for each verb, on each scale, namely "Pleasure", "Arousal" and "Dominance" respectively. Each verb is represented by an integer ID number in the interval [1, 25] (Table 1 shows the matching ID number-verb).
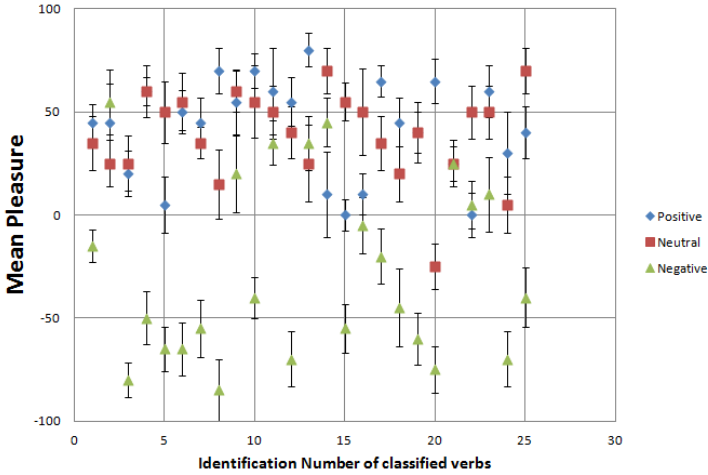

Fig. 3. The graphical representation of the average rating (with the corresponding standard error) of the 75 classified verbs on "Pleasure" dimension in percentage scale. Positive values correspond to positively assessed emotions, while negative values correspond to negatively assessed emotions.
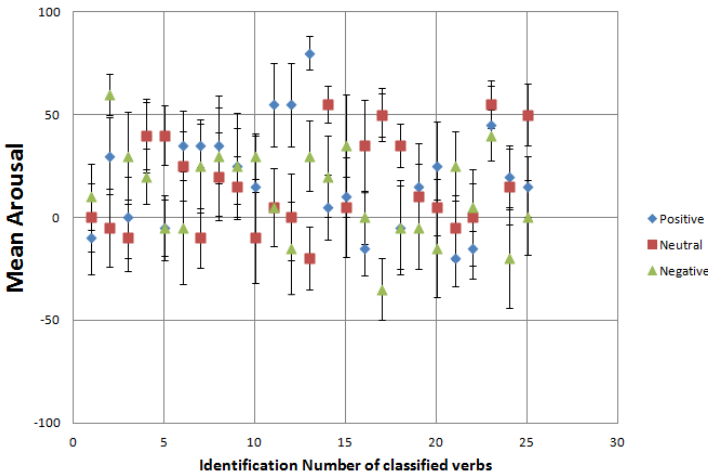

Fig. 4. The graphical representation of the average rating (with the corresponding standard error) of the 75 classified verbs on "Arousal" dimension in percentage scale. Positive values indicate higher intensity.
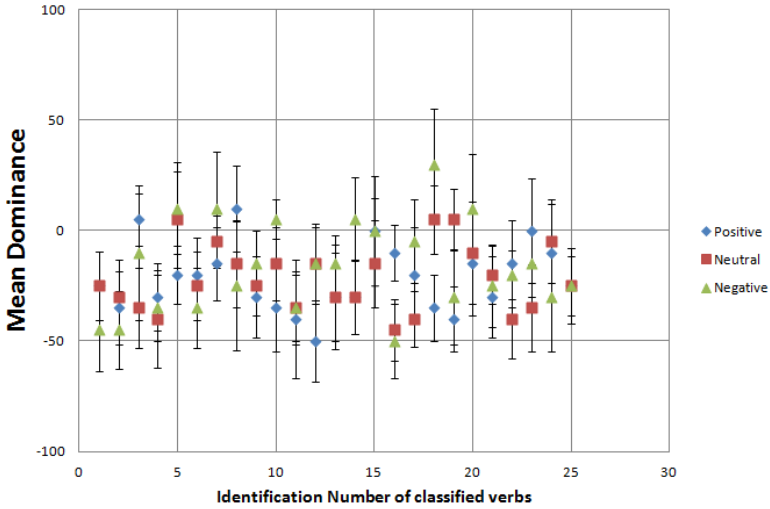
Fig. 5. The graphical representation of the average rating (with the corresponding standard error) of the 75 classified verbs on "Dominance" dimension in percentage scale. Positive values correspond to submissive emotions, while negative values correspond to dominant emotions.

Overall, positive verbs display a mean pleasure value of +42, contrary to negative verbs that have a mean value of -26.6. Neutral verbs also scored relatively high (mean 39), due to the positive connotations of verbs such as "can" (id 4), "know" (id 6), "find" (id 14) etc. There were no major differences across verb categories in the case of the other two axes, where mean arousal was 20.8, 16 and 11.4 for positive, neutral and negative verbs respectively, while mean dominance was -21.8, -21.8 and -16.2 respectively.

5   Discussion

The results of the pilot study indicate coherence and consistency in the evaluation. All positively classified verbs in SentiWordNet were assigned positive values with only two verbs ("tend" and "lend") scoring a zero, thus neutral value. Similarly, the majority of negative verbs is concentrated below the middle of the Pleasure axis. It should be noted that 8 out of the 25 negatively classified verbs recorded a positive score. We argue that this is mainly due to the inconsistencies in SentiWordNet's classification algorithm (e.g. "protect" (id 2)).

With regards to arousal, there is a high concentration of verbs around the middle of the y axis. Pairwise comparisons indicate that the evaluation was in the right direction. The verb "love" for example scored +51, while "prefer" scored merely -10. Similarly, in the case of dominance "fear", for example, which is submissive, scored +10, while "murder" and "save" scored -25 and -50 respectively. Brandley and Lang [24] present the results of the emotional rating, based on SAM test, called Affective Norms for English Words (ANEW). Our study extends ANEW in order to: a)

implement a methodology that can be accessed by a large statistical sample of participants and applied in different languages and cultures (web based experiment implementation), b) potentially ultimately serve as a crowdsourcing technique for fast and cost effective creation of lexical resources, c) provide insight into how people respond emotionally to the words they use, listen and/or read, as well as the affective meaning of the verb they perceive, d) include psychometric questionnaires (EPQ and SCL-90-R) and e) verify the validity of SentiWordNet respective results as a comparison to the ones of our own study.

The verbs are annotated using "Pleasure", "Arousal" and "Dominance" dimensions providing additional information than those in SentiWordNet. The addition of "Arousal" and "Dominance" dimensions overcomes this issue by offering enhanced information and showing the different influences (e.g. intensity, activation, potency).

Regarding the issues faced during the development of the whole experimental process, some users showed lack of concentration after the first half of the experiment, which suggests that the large number of verbs could be potentially decreased. Finally, we plan to add support of touch sensitive controls in order to make the experimental process easier and more straightforward, as well as support for blind users.

References

1. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Foundations and Trends, Information Retrieval 2, 1–135 (2008)
2. SentiWordNet, http://sentiwordnet.isti.cnr.it/
3. Esuli, A., Sebastiani, F.: SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In: 5th Conference on Language Resources and Evaluation (LREC 2006), Genoa, Italy, pp. 417–422 (2006)
4. Esuli, A., Baccianella, S., Sebastiani, F.: SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In: 7th Conference on International Language Resources and Evaluation (LREC 2010), Malta (2010)
5. WordNet, Lexical database of English, http://wordnet.princeton.edu/
6. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International Affective Picture System (IAPS): Affective retings of pictures and instruction manual. Technical report A-8, University of Florida, Gainesville, FL (2005)
7. Bradley, M.M., Lang, P.J.: Measuring emotion: The self-assessment manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry 25(1), 49–59 (1994)
8. Morris, J.D.: Observations SAM: The Self-Assessment Manikin - An Efficient Cross-Cultural Measurement of Emotional Response. Journal of Advertising Research 35(8), 63–68 (1995)

 9. Kouroupetroglou, G., Papatheodorou, N., Tsonos, D.: Design and Development Methodology for the Emotional State Estimation of Verbs. In: Holzinger, A., Ziefle, M., Hitz, M., Debevc, M. (eds.) SouthCHI 2013. LNCS, vol. 7946, pp. 1–15. Springer, Heidelberg (2013)
10. Alm, C.O., Roth, D., Sproat, R.: Emotions from text: Machine learning for text-based emotion prediction. In: Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP 2005), Vancouver, pp. 579–586 (2005)
11. Word frequency lists and dictionary from the Corpus of Contemporary American English, `http://www.wordfrequency.info/intro.asp`
12. Babylon 9 translation Software and Dictionary Tool, `http://www.babylon.com`
13. Scherer, K.R.: What are emotions? And how can they be measured? Social Science Information 44(4), 695–729 (2005)
14. Giouli, V., Fotopoulou, A.: Emotion verbs in Greek. From Lexicon-Grammar tables to multi-purpose syntactic and semantic lexica. In: 15th Euralex International Congress (EURALEX 2012), Oslo (2012)
15. Belletti, A., Rizzi, L.: Psych-Verbs and θ–theory. Natural Language and Linguistic Theory 6, 291–352 (1988)
16. Russell, J.A., Mehrabian, A.: Evidence for a three-factor theory of emotions. Journal of Research in Personality 11(3), 273–294 (1977)
17. PHP Hypertext Preprocessor, `http://www.php.net/`
18. Welling, L., Thomson, L.: PHP and MySQL Web Development, 4th edn. Addison-Wesley Professional, USA (2008)
19. HTML Tutorial, `http://www.w3schools.com/html/default.asp`
20. JavaScript Tutorial, `http://www.w3schools.com/js/default.asp`
21. CSS Reference, `http://www.w3schools.com/cssref/default.asp`
22. Sybil, B.G., Eysenck, H., Eysenck, J.: Paul Barrett: A revised version of the psychoticism scale. Personality and Individual Differences 6(1), 21–29 (1985)
23. Derogatis, L.R.: SCL-90-R: Administration, Scoring and Procedures Manual-II. Clinical Psychometric Research, Inc., Towson (1977)
24. Bradley, M.M., Lang, P.J.: Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida (1999)